NEW ZEALAND QUALIFICATIONS AUTHORITY
MANA TOHU MĀTAURANGA O AOTEAROA

ACHIEVEMENT STANDARD 90645

BI-VARIATE DATA ANALYSIS

NOTES FOR TEACHERS

This document has been prepared to provide some background information relevant to, and to assist teachers interpret the requirements of, the achievement standard 90645

ACHIEVEMENT STANDARD 90645 BI-VARIATE DATA ANALYSIS

NOTES FOR TEACHERS

This document has been prepared to provide some background information relevant to, and to assist teachers interpret the requirements of, the achievement standard 90645 *Select and analyse continuous bi-variate data*.

GENERAL

The standard is about the (possible) relationships between pairs of variables – specifically the strength of the linear relationship (correlation) and using any relationship to make predictions (regression). The purpose of the work done should be about investigating if any relationship exists between the variables and making predictions from a regression model for any relationship that is identified.

The standard specifies that the variables must be continuous. While regression analysis may be preformed on, and correlation coefficients calculated for, discrete variables, this requires care; using continuous variables avoids potential difficulties. Note that a discrete variable could be considered to be continuous in some circumstances (if it takes a large number of values) and a continuous variable could be considered to be discrete in some circumstances (if it has been recorded in a manner such that it takes a small number of values).

Students must *select* the variables to be investigated.

The analysis should primarily use a linear regression model, although other models could be considered for Merit and Excellence.

Some background information about the dataset should be given to students to enable the assumptions made, potential sources of bias and relevance and usefulness of evidence (aspects of the standard listed for Excellence) to be discussed without their having to resort to uninformed speculation. Comments made by students should be consistent with the background information.

If any of the possible aspects for Excellence that are listed in the standard are not relevant to a particular situation, reference to them should be omitted from the assessment task.

Comments made by students for Excellence need to be justified – eg for any proposed improvements to a model, an explanation should be given why/how the proposed model is an improvement. Comments should also be made in context, i.e. they should not be in general terms but should relate to the specific variables being investigated.

Before embarking on any analysis, students need to obtain an understanding of the data – which is another reason for providing some background information in the assessment task. They need to think about the data, how it may have been collected (or how it was collected if they did this themselves) and what the purpose of collecting the data may have been or was. Gaining an understanding of these aspects is helped by trying to visualise the data collection process.

It is important for students to clearly understand the role of the explanatory/predictor variable and the response/predicted variable and to clearly distinguish them before starting any analysis. The explanatory/predictor variable is plotted on the horizontal axis of a scattergraph and the response/predicted variable on the vertical axis (in considering correlation only, the

variables may be plotted on either axis). When regressing *y* on *x*, a least squares regression line (which is the most commonly used form of regression) minimises the sum of the squares of the *vertical* distances of the points from the regression line (i.e. the residuals). Hence, it is only valid to estimate (predict) *y* (the response/predicted variable) from *x* (the explanatory/ predictor variable). It is not valid to predict *x* from *y* – this would require *x* to be regressed on *y* and would result in a different regression equation.

If data are provided or if students collect their own data, any variables that are not relevant to the investigation need to be controlled – otherwise any observed effect may be due to one or more of these variables. For example, if the dataset has three variables *x*, *y* and *z*, the values of *x* and *y* need to be determined with *z* being held fixed. Any relationship that is found between *x* and *y* is only valid for the (fixed) value of *z*. In all likelihood, a different relationship between *x* and *y* will hold for a different value of *z*. In situations where there is more than one explanatory variable, multiple (multi-variate) regression needs to be used for those variables that have an interaction. This concept is developed a little further in the section Comments on Multiple Regression on pages 12 and 13. **Note that for this standard it is not expected that students be familiar with or use multiple regression.**

The following headings relate to particular specifications of the achievement standard.

ACHIEVEMENT

1 Describing the relationship between at least one pair of variables in context

In some contexts, it may be appropriate to describe the relationship between two variables by interpreting the gradient of a linear regression line – e.g. the change, on average, in sales revenue that is predicted from unit increase in advertising expenditure. This is another reason for fitting a linear regression model to the data in the first instance. Any such interpretation needs to be done in context. However, if students do describe the relationship between variables in this way, it is expected that they will also describe the correlation between the variables.

A relationship between variables is commonly expressed using a correlation coefficient (although correlation is specified in the standard for Merit). In describing a relationship, the strength of the relationship (through the magnitude of the correlation coefficient) and the direction of the relationship (through the sign of the correlation coefficient) should be explained and how these aspects relate to the appearance of a scattergraph of the variables (the greater the magnitude of the correlation coefficient, the smaller the scatter of points about the regression line).

In describing the relationship, students need to make it clear that they understand that there is a *tendency* for changes in the values of one variable to be associated with changes in values of the other. If the correlation coefficient is positive, the values of one variable *tend to* increase as the values of the other increase, and vice versa. If the correlation coefficient is negative, the values of one variable *tend to* decrease as the values of the other increase, and vice versa.

In discussing the relationship between two variables, reference could also be made to any

- clusters of points
- outliers
- change in the spread of the response/predicted variable as the explanatory/ predictor variable increases.

MERIT

1 Comparing the relationship between more than one pair of variables

Two investigations need to be carried out. It is expected that one variable will be common to the two investigations. The relationship between the variables in the first investigation can be compared with the relationship between the variables in the second investigation using the appropriate correlation coefficients. This provides another opportunity for students to show their understanding of correlation coefficients.

2 Discussing the appropriateness of the model

Discussion could include the following:

- how well the model fits the data, based principally on a visual inspection of a scattergraph with the regression line drawn and considering the scatter of the points about the line - for an appropriate model there should be no pattern to the scatter such as a concentration of points above the line near its ends and a concentration of points below the line near its centre
- the extent of the scatter of the points about the regression line - for either interpolation or extrapolation, the closer the value of R^2 is to 1, the more reliable predictions that are made will be because of the closer fit of the regression line to the observations; the value of R^2 can be used to support the evaluation provided by the visual inspection above
- the nature of the scatter of the points about the regression line – the scatter should be random with no pattern such as an increase or decrease in the scatter as the values of the explanatory/predictor variable increase
- the need to assume conditions remain constant when extrapolating, and the greater the distance of the extrapolation outside the range of observed values of the explanatory/predictor variable the less the confidence that can be held about the predictions.

Students need to understand that if the data they have is from a sample, the values they obtain for the parameters (constants) in the regression model are estimates of the values being sought. A different sample will probably result in different values for the parameters, and hence a different regression equation, and a different value for R^2.

In selecting between two or more regression models, a decision about which model is the best fit to the data should also be based primarily on a visual inspection of graphs and regression lines or curves as above. The decision should not be based on an increased value of R^2.

See the next section and the section Notes on *R*-squared and *r* (pages 9 – 12) for an explanation and discussion of R^2.

3 Interpreting correlation coefficients, *r*, and coefficients of determination, R^2, when appropriate

There are different correlation coefficients (for different purposes). The most common all-purpose correlation coefficient, and the one that is generally used when "the correlation coefficient" is referred to unqualified, is the Pearson product-moment correlation coefficient.

There are no absolutes as to what constitutes a "strong" relationship – it depends on the context. In cases where there is a lot of natural variability in each of the variables, a correlation coefficient of, say, 0.6 may be considered to show a strong relationship, but in cases where there is limited natural variability in each of the variables, a correlation coefficient of 0.6 may be considered to show only a moderately strong relationship.

It needs to be noted carefully that a correlation coefficient measures the strength of the *linear* relationship between the variables. This is also a good reason for fitting a linear regression model to the data in the first instance.

For linear (and polynomial) regression models, **R^2 measures the proportion of the total variability in the response variable that is accounted for or is explained by the regression effect between the variables as modelled by the regression function to which it refers** (the regression effect is the tendency of one variable to increase/decrease as the other increases/decreases). For example, if regressing sales revenue (the predicted or response variable) on advertising expenditure (the predictor or explanatory variable) has $R^2 = 0.72$ then, for that regression model, 72% of the variability in sales revenue is explained (accounted for) by the regression model; the remaining 28% is due to random factors (or perhaps other variables that were not controlled during the data collection process). It needs to be emphasised that *explaining* the variability is not the same as *causing* the variability; to minimise the risk of confusion, it may be better to use the phraseology "R^2 measures the proportion of the total variability in the response variable that is **accounted for** by the regression model".

It is sometimes said that R^2 measures the proportion of the total variability in the response variable that is explained by or accounted for by the **variability in the explanatory variable.** There is no foundation for this – see the section Notes on *R*-squared and *r* (pages 9 – 12) for an explanation.

For non-linear regression models (other than polynomial models) the meaning of R^2 cannot be quantified in the same way as it can for linear and polynomial models as the proportion of the variability in the response variable that is accounted for by the model - see the section Notes on *R*-squared and *r* (pages 9 – 12) for an explanation. However, the value of R^2 is still useful in supporting judgements about the reliability of predictions made from such models – if the value of R^2 is high, the data are scattered closely about the regression line or curve and the model gives reliable predictions.

Some authors and graphics calculators use the symbols *R* for the correlation coefficient and/or r^2 for the coefficient of determination. Teachers should insist that students use only *r* and R^2 to avoid confusion.

4 Making predictions from regression equations (interpolation and/or extrapolation)

It is expected that students will make two predictions, and that these will be interpreted in context. Making two predictions provides the opportunity for students to compare the results (and means that they need not be penalised for making an arithmetic mistake).

It is not appropriate to test how well a model fits a set of observations by selecting individual data point(s) and comparing the observed value with the predicted value of the response variable for a particular value of the explanatory variable. This is like using an individual value of a variable to represent all values of the variable. Any experimental data most likely contains random variation, and the random variation in any individual value(s) chosen to test the model may be greater than that in the other

values, so the point(s) chosen may be atypical. R^2 provides a measure of how well the model fits since it is calculated using all values of the response variable (in the same way that the mean usually provides the best single value that represents all values of a variable since it is calculated using all values of the variable).

Teachers may wish to distinguish between the *validity* and the *reliability* of a prediction.

- Validity refers to the *process* by which the prediction has been obtained. Least squares regression models, which are the most common and which are generally those used by statistical packages, give valid estimates (provided the data have been obtained using well-designed procedures – for example, there is only one explanatory variable that is likely to be relevant).
- Reliability refers to *how good* a prediction is. To be reliable, an estimate needs to be made from a model that is a good fit to the data.

5 Discussing the difference between correlation and causality when appropriate

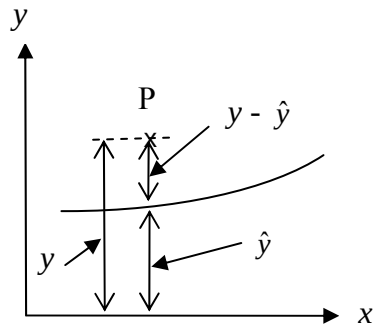
In some contexts, changes in the explanatory variable can be a direct cause of the changes in the response variable. An example of this is the amount of fuel used by vehicles (response variable) and the distance travelled by those vehicles (explanatory variable) – driving further is a direct cause of an increase in the amount of fuel used.

In some contexts, changes in the explanatory variable can be an indirect cause of the changes in the response variable – a third variable is involved (sometimes referred to as a confounding or "lurking" variable). An example of this is the relationship between infant mortality rates (response variable) and the number of registered medical practitioners per 100 000 people (explanatory variable) for different countries – one confounding variable is the per capita gross domestic product (others could be found as well). A change in the number of registered medical practitioners per 100 000 people is not the cause of a change in infant death rate! However, countries with a high per capita gross domestic product tend to have more registered medical practitioners per 100 000 people and lower infant mortality rates, and vice versa. Infant mortality rates and the number of registered medical practitioners per 100 000 people are both related to per capita gross domestic product, and as a result they are related indirectly to one another.

If a relationship is causal ("cause and effect"), this cannot be determined statistically – an understanding of the nature of the variables in the investigation is required. While some general knowledge may help (such as in the fuel usage example above) specialist knowledge may be needed before drawing conclusions about this aspect of the nature of any relationship.

6 Use of residuals

A residual (or prediction error) is defined as $e = y - \hat{y}$, where $\hat{y}$ (read "y-hat") is the estimated value of y obtained from the regression equation. Each observation (y value) has a residual. For y values larger than $\hat{y}$, the residual is positive; for y values smaller than $\hat{y}$, the residual is negative. With a least squares regression model, there will always be some positive residuals and some negative residuals. The diagram on the next page shows the residual for the observation represented by the point P for the (non-linear) regression model shown. Since P is above the line its residual is positive.



Residuals may be obtained automatically in Excel using the regression tool (if installed) from the TOOLS menu or by using a graphics calculator. They may also be readily calculated manually in Excel.

A residual plot graphs the residuals on the vertical axis against the values of the predictor/explanatory variable on the horizontal axis (other forms of residual plot are sometimes used in more advanced statistical analysis). While a scattergraph also shows the residuals, a residual plot provides a more direct view of the residuals since the regression line becomes the horizontal axis.

For reliable regression estimates, the residuals should have

- limited variability
- no pattern such as generally positive residuals at one end of the graph and generally negative residuals at the other end
- constant variability as the value of the explanatory variable changes – a reasonable number of observations is required to make judgements about this.

A least squares linear or polynomial regression model is fitted so that it minimises the sum of the squares of the residuals (which gives rise to the term least squares). Power and exponential models are found by fitting the least squares regression line to the linearised variables (by making log/log and log/linear transformations of the variables respectively).

EXCELLENCE

1 Methods of analysis

This is possibly only relevant if unusual methods or concepts beyond the expectations of the standard are used.

2 Assumptions made

An understanding of the data is required for this. If students do not have an understanding of the data, they are likely to resort to uninformed speculation.

Any comments made need to be realistic and related to the context.

3 Limitations

Limitations must refer to the analysis and not to aspects such as the experiment or the data collection process.

Consideration should be given to what happens at the extremes of the fitted model. For example:

- when $x = 0$ the model may predict a value of y that is not 0 but this presents logical contradiction when the nature of the variables is considered
- as x gets larger indefinitely the model may predict that y also increases indefinitely, but consideration of the nature of the variables may suggest a different behaviour (e.g. the value of y may taper off or there may be a physical limit to the value of y)
- a line with a negative gradient will cross the x-axis for some value of x, and for larger values of x it will result in the prediction of negative values of y – these may have no physical meaning
- there may be physical limitations of the variables – e.g. the heights of adults have upper and lower bounds.

Note that the vertical intercept of the regression line can be set at a specified value using Excel. In particular, the nature of the variables may determine that $y = 0$ when $x = 0$, and a regression model with this feature can be found (in particular, Excel allows the y-intercept to be set by the user (under the menu option "Format trendline" and then "Options", there is an option to set the intercept).

Discussion could include how well the model fits the raw data over its entire range.

4 Improving regression models eg discussing the effect of outliers, fitting piecewise or non-linear models

Non-linear or piece-wise models could be considered. Models in which the y-intercept is set to 0 because of considerations about the nature of the variables may also be considered.

While there are formal tests to determine if points should be classified as outliers, it is not expected that such tests will be used at this level. Outliers should be quite distinct from other points in the scattergraph, and not just on the periphery. If any point in a scattergraph is identified as an outlier, the first action taken should be to try and find a possible explanation for it - an outlier could be the result of a data entry mistake for instance, or there may be some identifiable possible (realistic) or actual cause. Outliers must be treated very carefully, and any action taken fully justified – e.g. because it is a "one-off" observation; any possible cause that is suggested needs to be reasonable and not simply uninformed speculation. Removal of an outlier from the analysis should not be justified purely on the grounds that it improves the fit of the model (and increases the value of R^2).

Any reference to “having more data” to improve the analysis needs to be fully justified. Having more data may not actually improve the situation as there may be more variability in one or both variables with more observations.

5 Alternative approaches

This aspect will not always be relevant. Students may describe another approach that could be taken to the analysis, but any discussion needs to be realistic and not just speculative.

6 Data source or data collection method if the student collects own data

This is only relevant in special cases.

7 Potential sources of bias

Bias is the distortion of the outcomes of sampling that results from a systematic action or actions taken regarding the selection of sample members. For example, if a survey is to conducted in a city, collecting data from particular areas of the city may lead to biased results.

Bias can be expected to be minimal or to have no effect in any well-designed experiment.

If students conduct an experiment or a survey to obtain their data, they need to think about possible sources of bias before starting, and take steps to minimise their effects.

8 Relevance and usefulness of evidence

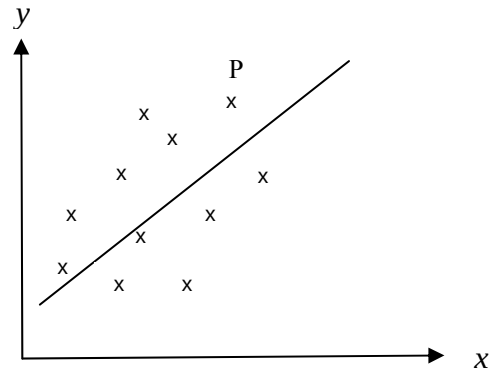
Discussion should be about who would make use of the results of the analysis and for what purpose, and the extent to which the data and any prediction(s) would enable them to do this. Any such comments need to be realistic.

9 How widely the findings can be applied

A possible target population(s) needs to be inferred. For example, suppose that data have been obtained from some species of mammals randomly sampled from a particular region. Then the findings can be applied to all mammals of those species in that region. If, however, it can be safely assumed or there is evidence that the characteristic(s) investigated for these species is/are no different from the same characteristic(s) in these species in other regions, then the findings can be also applied to the same species in those regions. Similarly, if it can be safely assumed or there is evidence that the characteristic(s) investigated is/are no different from the same characteristic(s) for other species, then the findings can also be applied to those other species.

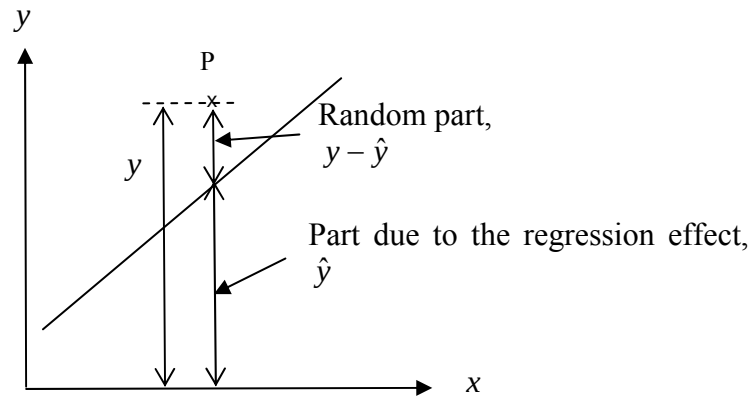
NOTES ON *R*-SQUARED AND *r*

Consider the scattergraph of two variables *x* and *y* shown below. One specific point, P, has been marked. The scattergraph shows that there is a regression effect between *x* and *y* (a regression effect is the tendency for one variable to increase or decrease with an increase in the other, in this case for *y* to increase as *x* increases). A regression line has also been drawn on the scattergraph.

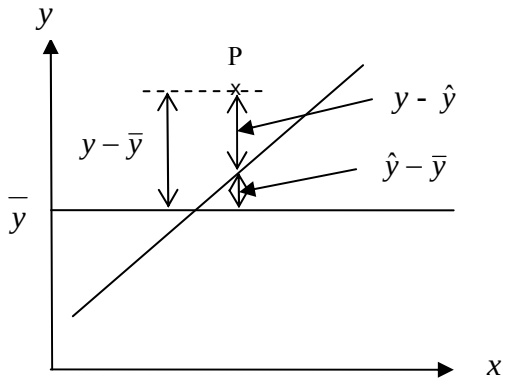


The regression effect is not perfect - if it was, the points would all lie on the regression line. The variation in *y* about the regression line is not constant but is random. Hence any value of *y* has two parts – a part due to the regression effect and the remainder a random effect.

The graph below shows the same scattergraph but with just the point P shown. The two components that make up the *y* value of P are also shown. $\hat{y}$ ("*y*-hat") is the regression estimate of *y* (the value of *y* as predicted by the regression model).



If there was no regression effect and no random effects, then all *y* values would be the same, i.e. equal to $\bar{y}$ (the mean value of *y*), so it is natural to consider the situation above with $\bar{y}$ as a baseline or reference value. The graph below is the same graph as above but with the deviations of *y* and $\hat{y}$ from $\bar{y}$ shown.



For linear regression models, it can be shown that

$$\sum (y - \bar{y})^2 = \sum (y - \hat{y})^2 + \sum (\hat{y} - \bar{y})^2$$

$\sum (y - \bar{y})^2$ is the total sum of squares (SST). It is a measure of the total variability of the variable *y* about its mean. Note that the mean of this value, $\frac{\sum (y - \bar{y})^2}{n}$, is the variance of *y*.

$\sum (y - \hat{y})^2$ is the sum of squares due to prediction errors (SSE). It is a measure of the variability of *y* due to random or irregular effects (the errors) - *the unexplained variation*. Note that each $y - \hat{y}$ is a residual. SSE is sometimes referred to as the sum of squares due to residuals.

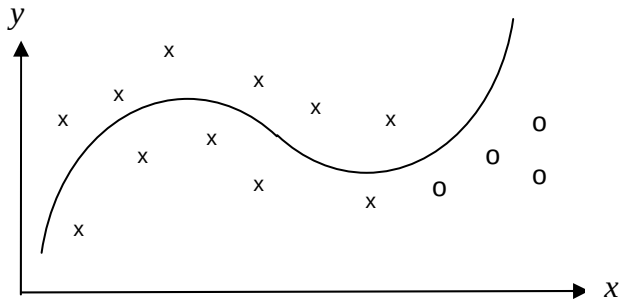
$\sum (\hat{y} - \bar{y})^2$ is the sum of squares due to the regression effect (SSR). It is a measure of the variability of *y* due to or accounted for by the regression effect – *the explained variation*.

R^2 is defined as follows: $R^2 = \dfrac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2} \left(= \dfrac{SSR}{SST}\right)$. Hence, R^2 is the ratio of the explained variability to total variability, i.e. **R^2 is the proportion of the total variability in *y* that is accounted for or explained by the regression effect as modelled by the regression function to which it refers** (note that a different regression function for the same data would have a different value of R^2)**.** The remaining variability is due to random effects.

Note that since R^2 is a ratio of two sums of squares, then it must be positive. Further, since $\sum(\hat{y}-\bar{y})^2 \leq \sum(y-\bar{y})^2$ (with equality only in the case of perfect regression) then $R^2 \leq 1$. Hence, $0 \leq R^2 \leq 1$. Also, since R^2 is a ratio, it has no units.

The closer the value of R^2 is to 1, the greater the proportion of the total variability in *y* is explained and the smaller the proportion of the total variability in *y* that is due to random effects, so the closer the points are to the regression line. Hence the closer R^2 is to 1, the better the regression model is a fit to the data.

The identity $\sum(y-\bar{y})^2 = \sum(y-\hat{y})^2 + \sum(\hat{y}-\bar{y})^2$ also holds for polynomial regression models (and so the definition and interpretation of R^2 also apply to polynomial models). While such models may be useful for interpolating, they may be quite misleading when extrapolating as the following diagram shows. The cubic polynomial is a good fit to the data plotted as crosses. However, the extended data plotted as circles show that the curve does not fit the extended data set well. Using the cubic to make predictions in the range where the data are shown as circles or beyond would clearly produce non-meaningful results.



Polynomial regression models need to be used with considerable care. Teachers may wish to prevent difficulties by avoiding them.

The identity $\sum(y-\bar{y})^2 = \sum(y-\hat{y})^2 + \sum(\hat{y}-\bar{y})^2$ does not apply to non-linear regression models such as power and exponential models. For these models the total variability in *y* cannot be split into two components. As a result the ratio $\dfrac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2}$ does not have the same meaning as it does for linear (and polynomial) regression models. For such models, R^2 is calculated from a transformation of the variables (eg a log/log transformation) in which the variables are linearised. For these models, R^2 cannot be interpreted in the same way as it can for linear models - it applies to the transformed variables. Hence, comparisons of R^2 values should not be made between linear and non-linear models or between different forms of non-linear models. However, it is still true for non-linear models that the closer the value of R^2 is to 1 the better the regression model fits the data and vice versa.

As noted previously, it is sometimes said that R^2 measures the proportion of the total variability in the response variable that is explained by or accounted for by the *variability in the*

explanatory variable. However, the definition of R^2 makes no reference to x, and so the statement is incorrect. To illustrate this further, consider the tables below. The right hand table contains the same data as the left hand table except that the values of x have been multiplied by 3. In the table, SST is $\sum(x-\bar{x})^2$, which is the total variability in x. The variance of x has also been shown, showing that it is SST/10 as expected ($n = 10$). Although the variability in x is greater in the right hand table than it is in the left hand table, the R^2 value has not changed. The variability in x has no effect on the value of R^2.

x	*y*
1.3	4.1
2.6	6.4
2.5	3.3
3.6	5.5
3.4	8.9
3.0	6.0
3.5	6.6
3.9	7.7
5.1	7.9
5.3	10.8

SST =	12.816
Var(X) =	1.2816
R^2 =	0.6417

x	*y*
3.9	4.1
7.8	6.4
7.5	3.3
10.8	5.5
10.2	8.9
9.0	6.0
10.5	6.6
11.7	7.7
15.3	7.9
15.9	10.8

SST =	115.344
Var(X) =	11.5344
R^2 =	0.6417

For a linear regression model, the (Pearson product-moment) correlation coefficient r is defined as $\pm\sqrt{R^2}$, with the sign taken being that of the gradient of the regression line. Hence $-1 \le r \le 1$. If there is a tendency for the values of one variable to increase as the other increases (and the regression line therefore has a positive gradient) then r is positive, and vice versa. Note that r is a measure of *linear* association.

R^2 and r are closely related. However, taking the square root gives a measure (r) that is associated directly with the units in which measurements have been made rather than the squares of the units (the variance and standard deviation of a variable are associated in a similar way). It needs to be noted, though, that being based on a ratio, neither r nor R^2 have units.

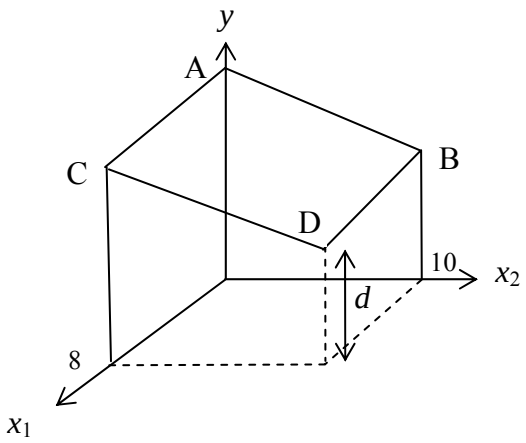
COMMENTS ON MULTIPLE REGRESSION

Multiple (multi-variate) regression is used in situations where there is more than one explanatory variable and the simultaneous effect of these is to be considered. For example, the yield of a particular crop may depend on a number of variables such as the soil moisture, the soil temperature, the pH of the soil and the amount of fertiliser applied. Examples of multi-variate regression equations are:

- the equation $y = 2x_1 + 5x_2 + 3$ represents a situation in which a variable y, the response variable, is regressed on two explanatory variables x_1 and x_2 - this is a linear regression model since it involves a linear function of each variable in the equation
- the equation $y = 2x_1^3 + 5x_2 - 4x_3^2 - 7$ represents a situation in which a variable y, the response variable, is regressed on three explanatory variables x_1, x_2 and x_3 - this is a non-linear regression model since it involves non-linear functions of at least one explanatory variable, in this case, x_1 and x_3.

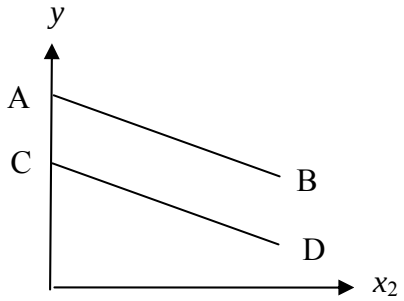
Note that, in practice, the coefficients in a regression equation are most likely to be non-integral.

A regression of two variables can be visualised as a surface in three dimensions. If the regression model is linear, the surface is a plane. If the regression model is non-linear, the surface is curved. The following diagram represents a linear regression model.



Any height of the surface above the x_1-x_2 plane represents a regression estimate. The height d shown in the diagram represents the regression estimate of y for $x_1 = 8$ and $x_2 = 10$.

If x_1 is held constant, the diagram above degenerates into a plane parallel to the y-x_2 plane through that value of x_1, and the regression surface degenerates into a line. AB represents the regression model for the case where $x_1 = 0$, and CD represents the regression model for the case where $x_1 = 8$. The following diagram shows the situation. Since AB and CD are not coincident, they will have different equations



Similarly if x_2 is held constant, the diagram above degenerates into a plane parallel to the y-x_1 plane through that value of x_2.

The effect of holding the value of one variable constant can also be demonstrated analytically. Suppose that $y = 12 - 2x_1 - 5x_2$ is the model for regressing y on x_1 and x_2. If x_1 is held constant at, say, 0, the relationship between y and x_2 is represented by $y = 12 – 2 \times 0 - 5x_2$, i.e. $y = 12 - 5x_2$. However, if x_1 is held constant at, say, 1, then the relationship between y and x_1 is represented by $y = 12 – 2 \times 1 - 5x_1$, i.e. $y = 10 - 5x_2$, resulting in a different regression equation.

Hence, if more than one explanatory variable may affect the response variable but multiple regression analysis is not to be used (so that the investigation is to involve just two variables) then all other variables must be held constant, and the regression model holds for just the constant value(s) of the other variable(s).

As noted previously, it is not expected for this standard that students be familiar with or use multiple regression.